FG117 (1)

LOAD BALANCER

Background of the Invention

【0001】

Field of the Invention

The present invention relates to a load balancer, and in particular to a load balancer for relieving concentrations of a server load.

【0002】

Description of the Related Art

Various forms of load balancers have been known. An example thereof is shown in Fig.16 where a load balancer LB receives a packet transmitted from a client CL through routers R1 and R3, and selects one of servers as a destination to be connected according to a source address (SA) included in an IP header of the packet or to a combination of the source address and a TCP port number.

【0003】

Also, there is a form of designating a server distribution destination on an upper application level, or a system of selecting a server according to a URL of the destination to be connected for example.

Moreover, server distributions have been made possible by identifying an SSL (Secure Sockets Layer; an encryption protocol for ensuring TCP/IP communication security on a transport layer) session ID or a Cookie variable for identifying a user where a WWW server generates character string information "Cookie variable" to be stored by both of a server and a browser.

【0004】

Another significant point as to functions of such a load balancer

is a function of server consistency (transaction consistency) during communications.

For example, when server loads used for online shopping are balanced, a "consistency maintenance function" is essential for maintaining a correspondence between the client and the server during a series of flows from a product selection to a purchase and a settlement thereof.

【0005】

In case of a load balancer unprovided with such a consistency maintenance function, accesses are distributed to different servers during the communication, so that a series of processing of an electronic commerce can not be operated properly.

Therefore, many of the load balancers are provided with the consistency maintenance function based on IP header information.

【0006】

While the client CL shown in Fig.16 is a fixed terminal, a mobile IP technology for managing movements of terminals on an IP protocol is currently proposed by the IETF (Internet Engineering Task Force). This technology is for maintaining a communication even if an IP terminal moves, and realizes the mobile communication by acquiring a new care-of address (CoA) whenever the IP terminal moves and by registering the care-of address in a home agent (HA) or a destination terminal.

【0007】

When a server is accessed from an IP terminal utilizing this technology and when a server load balancing is performed, an IP address (care-of address) of the IP terminal is changed whenever it moves, so that the source address is changed. Therefore, the function of server consistency maintenance during communications using the source address (or combined with the source address) cannot be accomplished.

【0008】

This will be described referring to a schematic diagram of a packet transmission when a mobile IPv6 is utilized as shown in Fig.1.

A packet whose source address (SA) is a care-of address CoA1 when a mobile node (hereinafter, referred to as mobile IP terminal) MN is connected to a router R1 belonging the area or a care-of address CoA2 when connected to a router R2 belonging the area by the movement is prepared and transmitted.

【0009】

Assuming that the load balancer LB is provided between the router R3 and servers S1, S2 in the same way as in Fig.16, when the load balancer LB receives such a packet and uses the source address (SA) as a retrieval (selection) key in the same way as in Fig.16, the load balancer misidentifies an access from the same mobile IP terminal MN as an access from another mobile IP terminal, thereby changing the server S1 to S2, so that there has been a problem that the server consistency cannot be guaranteed.

【0010】

Summary of the Invention

It is accordingly an object of the present invention to provide a load balancer in which a function of server consistency maintenance during communications is preserved even if a terminal is a mobile IP terminal.

【0011】

In order to achieve the above mentioned object, a load balancer according to the present invention comprises: means for extracting identifying information specific to a mobile IP terminal from an arrival packet; and means for determining a destination server to be connected based on the identifying information. (claim 1)

【0012】

Namely, in the present invention, in order to provide a load balancing function to a mobile IP terminal, identifying information specific to a mobile IP terminal is extracted from an arrival packet. The packet is distributed to the destination server to be connected based on the extracted identifying information.

For the above-mentioned identifying information, a home address included in a destination option header added to an IP packet transmitted by the mobile IP terminal while moving may be used. (claim 2)

【0013】

Namely, as shown by an example of a server-addressed packet format in case a mobile IPv6 is used (see Fig.4), a home address is the same while a care-of address (source address) of a packet received by the load balancer is changed with the movement of the mobile IP terminal. Therefore, by using this home address for the server selection, maintenance of a server connection consistency can be achieved.

【0014】

Also, for the above-mentioned identifying information, what is prescribed in predetermined lower bits of a source address of a packet utilizing a stateless address configuration method may be used. (claim 3)

Namely, there are two preparation systems (methods) of a mobile IPv6 address as follows: (1) stateless address generation system (stateless auto address configuration) where a terminal freely prepares an address to be used, and (2) stateful address generation system (stateful auto address configuration) where a terminal obtains permission when using an address. Among these, when the above-mentioned stateless system (1) is used, L2 address identifier of the mobile IP terminal itself (e.g. the number of combined MAC

address in case of an Ethernet) is used for lower 64 bits (see Fig.6) of a care-of address prepared by the mobile IP terminal.

【0015】

Since the address identifier is a value specific to a mobile IP terminal, it becomes possible for the load balancer to select the server by using the lower 64 bits of the received packet or by combining the 64 bits with other information.

Moreover, for the above-mentioned identifying information, a security parameter index (SPI) of the packet if encrypted may be used. (claim 4)

【0016】

There are cases in which a mobile IPv6 packet is encrypted by an IPSEC (security) for security measures of the packet itself. In such a case, in areas following the encryption area, codes can not be decoded if there is no encryption key used between both ends.

【0017】

In such a case, a security parameter index (SPI: see Fig.9) can be used to guarantee a consistency of server selection. This security parameter index (SPI) is a number agreed upon by both ends in order to indicate the relationship between the encryption algorithm and the encryption key number used by both ends, and is written in the head portion of the encryption payload.

【0018】

Also, a load balancer according to the present invention may comprise: means for requesting a home agent to notify a change of a care-of address when the care-of address of a terminal has changed upon an arrival of a first packet addressed to a server; and means for determining a destination server to be connected by regarding the notified care-of address as identifying information. (claim 5)

【0019】

Namely, a care-of address of a receiving packet is changed as the

mobile IP terminal moves. The relationship between a care-of address and a home address is managed by a home agent. Therefore, when receiving a packet including a care-of address, the load balancer may instruct the home agent to always notify a new care-of address when the care-of address has been changed, so that even if no packet is generated during communications, the home agent notifies the changed care-of address to the load balancer when the terminal moves.

【0020】

Thus, the load balancer always knows the latest care-of address, and can select the destination server using same.

In the above-mentioned case, the home agent is adapted to notify the changed care-of address to the load balancer upon the change thereof. However, considering that the mobile IP terminal itself knows the change of the care-of address, the mobile node itself may be requested as a substitute for a home agent to notify the change of the care-of address when the care-of address in the first arrival packet has changed, and a server connection distribution may be performed by regarding the notified care-of address as identifying information. (claim 6)

【0021】

Also, when the above-mentioned extracting means extract a packet transmitted from a home link upon an arrival of the packet and the packet does not have the destination option header, the determining means may determine the destination server by regarding a source address of the packet as the identifying information. (claim 7)

【0022】

Namely, when the first packet arrives and if the packet is transmitted from a home link, or transmitted from an unmoved mobile IP terminal, the source address of the packet is regarded as the

identifying information to determine the destination server.

【0023】

Also, the above-mentioned determining means may be provided with a table for storing an address of the destination server having a source address of a packet, i.e. the care-of address, as a retrieval key, thereby determining the destination server using the source address of the arrival packet. (claim 8)

【0024】

Also, the above-mentioned determining means may be provided with a table for storing an address of the destination server having a source address of a packet, i.e. the care-of address, as a retrieval key, thereby determining the destination server using the source address of the arrival packet, and the table may prepare an entry with a new care-of address as a retrieval key when the new care-of address has been notified, and may store, as storing data, an address of the destination server stored as data of an entry of an old care-of address. (claim 9)

【0025】

Also, the above-mentioned determining means may store a lifetime in the data of the entry, may periodically decrement the lifetime, may update the lifetime every time a packet using the entry has arrived, and may invalidate the entry upon expiration of the lifetime. (claim 10)

【0026】

Also, a home agent of a mobile IP terminal as a substitute for the above-mentioned server may be made a destination to be connected. (claim 11)

Namely, not only a server but also a home agent of a mobile IP terminal may be made an object of a load balancing.

【0027】

Moreover, a home agent may be realized which notifies,

according to a request from a load balancer, binding cache information managed by the home agent itself to the load balancer periodically or when triggered in operation by a change of a care-of address of a mobile IP terminal. (claim 12).

【0028】

Moreover, a mobile IP terminal may be realized which notifies, according to a request from a load balancer, binding cache information managed by the mobile IP terminal itself to the load balancer periodically or when triggered in operation by a change of a care-of address of the mobile IP terminal itself. (claim 13).

Brief Description of the Drawings

Fig.1 is a diagram showing a network embodiment (1) to which a load balancer according to the present invention is applied;

Fig.2 is a block diagram showing an arrangement (1) of a load balancer according to the present invention;

Fig.3 is a flow chart showing an operation example (where a home address is made a retrieval key) of an IPv6 header information extractor used in an arrangement (1) of a load balancer according to the present invention;

Fig.4 is a format diagram showing a mobile IPv6 packet (from terminal to server) used in an arrangement (1) of a load balancer according to the present invention;

Fig.5 is a diagram showing an arrangement (where a home address is made a retrieval key) of a load balancing table used in an arrangement (1) of a load balancer according to the present invention;

Fig.6 is a format diagram of a general global address of an IPv6 packet;

Fig.7 is a diagram showing an arrangement (where lower 64 bits of a source address is made a retrieval key) of a load balancing table used in an arrangement (1) of a load balancer according to the present

invention;

Fig.8 is a flow chart showing an operation example (where an SPI is made a retrieval key) of an IPv6 header information extractor used in an arrangement (1) of a load balancer according to the present invention;

Fig.9 is a format diagram showing an IPSEC-encrypted mobile IPv6 packet (from terminal to server) used in an arrangement (1) of a load balancer according to the present invention;

Fig.10 is a diagram showing an arrangement (where an SPI is made a retrieval key) of a load balancing table used in an arrangement (1) of a load balancer according to the present invention;

Fig.11 is a diagram showing a network embodiment (2) (when a first packet is received) to which a load balancer according to the present invention is applied;

Fig.12 is a diagram showing a network embodiment (2) (when a mobile IP terminal moves) to which a load balancer according to the present invention is applied;

Fig.13 is a block diagram showing an arrangement (2) of a load balancer according to the present invention;

Figs.14A and 14B are flow charts showing an operation example (where a source address is made a retrieval key) of an IPv6 header information extractor used in an arrangement (2) of a load balancer according to the present invention;

Fig.15 is a diagram showing an arrangement (where a source address is made a retrieval key) of a load balancing table used in an arrangement (2) of a load balancer according to the present invention; and

Fig.16 is a diagram showing a network arrangement to which a prior art load balancer is applied.

Throughout the figures, like reference numerals indicate like or corresponding components.

【0029】

Description of the Embodiments

Network embodiment (1)

Fig.1 shows a network embodiment (1) to which a load balancer according to the present invention is applied. In this embodiment, when a mobile IP terminal MN moves from a router R1 belonging the area to a router R2 belonging the area and when its care-of address is changed from CoA1 to CoA2, identifying information (retrieval key) specific to the mobile IP terminal MN is extracted, so that if a server S1 has been selected the server S1 is selected similarly during communications, thereby retaining the function of server consistency maintenance.

【0030】

Fig.2 shows an arrangement (1) of the load balancer LB shown in Fig.1. In this arrangement, the load balancer LB is composed of an IPv6 header information extractor 1, a load balancing table 2, a table controller 3, and a distribution processor 4.

The IPv6 header information extractor 1 extracts identifying information, so that the table controller 3 looks up the load balancing table 2 regarding the identifying information as a retrieval key. The distribution processor 4 determines a destination server to be distributed based on the server information thus obtained.

【0031】

It is to be noted that the load balancing table 2, the table controller 3, and the distribution processor 4 form the determining means.

(1) In case home address is used as retrieval key:

Fig.3 shows an operation example of the IPv6 header information extractor 1 in case a home address is used as the above-mentioned retrieval key (identifying information) to select a

server based thereon.

【0032】

Namely, a header extraction is firstly executed (at step T1), where a "Next header" field is checked sequentially from an IPv6 header 10 of a mobile IPv6 packet (from terminal to server) shown in Fig.4 until a code "60" indicating an IPv6 destination option is found (at step T2).

【0033】

When the destination option code "60" is not found, it is determined that the mobile IP terminal MN has not moved but been connected to a home link, so that the table controller 3 retrieves a destination server to be connected from the table 2 using a source address (SA) as the retrieval key, as later described. The distribution processor 4 transmits the packet to the server (at step T3).

【0034】

On the other hand, when the destination option code "60" is found, it is indicated that a destination option header 20 exists in the field following the IPv6 header.

Subsequently, it is determined whether or not an option type included in the destination option header 20 is "201" (at step T4). If the option type is not "201", the routine proceeds to the above-mentioned step T3. If the option type is "201", this option type indicates a home address option, namely, the mobile IP terminal MN has moved. Therefore, the distribution processor 4 selects a corresponding server from a server group using the home address as the retrieval key (at step T5).

【0035】

Fig.5 shows a table arrangement of the load balancing table 2 shown in Fig.2. In this table, with a home address "2000.16" being made a retrieval key, this home address and a destination server address "S1" associated therewith one-to-one are stored.

【0036】

The destination server address is the address of the server determined according to a load balancing algorithm at an initiation of communications. There are various methods or algorithms for selecting the destination server such as a method selecting a server by a round robin, and a method selecting one with less server loads. These conventional methods can be used unchanged.

【0037】

It is to be noted that the home address in Fig.5 is shown having a format where "2000" corresponds to a network ID (64 bits) of a global address format of a mobile IPv6 packet shown in Fig.6, and "16" corresponds to an interface ID (64 bits). Hereinafter, the address indication format of the retrieval key is the same.

【0038】

Also, the retrieval table of the destination server address making the source address (SA) a retrieval key at step T3 of Fig.3, is similarly stored in the above-mentioned load balancing table 2 (reference numerals are also the same). This table example is the same as that of Fig.5 except that the source address SA in substitute for the home address is written in the field of the retrieval key.

<u>(2) In case lower 64 bits of SA address are used as retrieval key:</u>

Apart from the case where the home address is made the retrieval key as mentioned above, when the packet is prepared by the stateless address generation system, it is also possible to make the lower 64 bits of the global address of the mobile IPv6 shown in Fig.6 a retrieval key. It is to be noted that the arrangement (1) of Fig.2 can be used as the arrangement of the load balancer LB in this case.

【0039】

Namely, in case of the stateless address generation system described above, the lower 64 bits of the global address of the mobile IPv6 packet use an L2 address identifier such as a number obtained

by combining therewith e.g. the MAC address of the terminal itself as for an Ethernet.

【0040】

Therefore, since the address is unique to each terminal, the lower 64 bits of the global address may be extracted instead of searching for the home address.

Also, the arrangement of the load balancing table 2 in this case is shown in Fig.7. If the lower 64 bits of the source address (SA) are "16", it is indicated that the destination server address is "S1", so that the distribution processor 4 distributes the connecting destination of the server based on the retrieval result.

(3) In case SPI is used as retrieval key:

Apart from using the home address or the lower 64 bits of the source address as mentioned above, the load balancing can be performed using a security parameter index (SPI) when the mobile IPv6 packet is encrypted. It is to be noted that the arrangement (1) of Fig.2 can also be used for the arrangement of the load balancer LB in this case.

【0041】

An operation example of the IPv6 header information extractor 1 in this case is shown in Fig.8. Firstly, the IPv6 header information extractor 1 performs the header extraction as mentioned-above (at step T11). In this case, as shown in the IPSEC-encrypted format of the mobile IPv6 packet (from terminal to server) of Fig.9, the "next header" field is checked sequentially from the IPv6 header 10 until a code "50" indicating an encryption header (ESP header) 40 is found (at step T12).

【0042】

If the code "50" is not found, the server is selected based on the source address (SA) (at step T13), and otherwise the next option header is the ESP header 40.

At the top of the ESP header 40, the security parameter index SPI is written, so that the index SPI thus acquired is made the retrieval key of the table to be looked up next (at step T14).

【0043】

Fig.10 shows an arrangement of the load balancing table 2 when the index SPI is thus made the retrieval key. In this example, the destination server address "S1" associated one-to-one with the index SPI "218" is stored. The destination server address "S1" is the address of the server determined by the load distribution algorithm at the initiation of the communication as mentioned above.

【0044】

Network embodiment (2)

Figs.11 and 12 show a network embodiment (2) to which a load balancer according to the present invention is applied. This embodiment is a system where information is exchanged between the home agent HA and the load balancer LB so that the retrieval key is changed with the movement of the mobile IP terminal MN.

【0045】

Hereinafter, an overall operation of this embodiment will be described referring to Figs.11 and 12.

Firstly, the operation example when the load balancer LB receives the first packet from the mobile IP terminal MN as shown in Fig.11 will be described.

① The mobile IPv6 packet from the mobile IP terminal MN at a care-of address CoA1 is received by the load balancer LB through routers R1 and R3.

【0046】

② The load balancer LB extracts the source address (SA=CoA1) of the mobile IP terminal MN as well as the home address of the terminal MN.

③ According to the extracted care-of address CoA1 (and other

information), e.g. a server S1 is selected as the destination server. At the same time, a header rewrite process, etc. are executed, and the packet is transmitted to the selected server S1.

【0047】

④ With the above-mentioned ③, the load balancer LB transmits a care-of address notification request, for having the fact of the care-of address change notified when the care-of address is changed, from a CoA notification request generator 5 to the home agent HA accommodating the mobile IP terminal MN.

⑤ The home agent HA returns a response to the care-of address notification request.

【0048】

An operation example when the mobile IP terminal MN further moves as shown in Fig.12 will be described.

① The mobile IP terminal MN moves from the router R1 belonging the area of the care-of address CoA1 to the router R2 belonging the area of the care-of address CoA2.

【0049】

② The mobile IP terminal MN acquires the care-of address CoA2.

③ The mobile IP terminal MN transmits a binding update indicating that the care-of address has been changed from CoA1 to CoA2 to the home agent HA.

④ The home agent HA having received the binding update updates the care-of address in the binding cache of the home agent HA itself from CoA1 to CoA2.

【0050】

⑤ The home agent HA notifies the binding update of the care-of address CoA2 to the load balancer LB.

⑥ The load balancer LB updates the care-of address from CoA1 to CoA2 in its internal cache in the same way as the home agent HA.

【0051】

⑦ The load balancer LB then selects the server S1 based on the care-of address CoA2 in the same way as the care-of address CoA1.

An arrangement (2) of the load balancer LB executing such an operation is shown in Fig.13 which differs from that shown in Fig.2 in that the CoA notification request generator 5 has been added.

【0052】

Hereinafter, an operation example of the IPv6 header information extractor 1 shown in Figs.14A and 14B in the arrangement (2) of the load balancer LB will be described with some portions being overlapped with the above-mentioned description.

Firstly, the IPv6 header information extractor 1 extracts the header of the received packet as described above (at step T21 of Fig.14A). Usually, since the source address (SA) as the care-of address is used as the retrieval key, whether or not the received packet is a notification message ⑤ shown in Fig.12 indicating the change of the source address (SA) is determined (at step T22).

【0053】

As a result, if it is not the message in which the source address (SA) is changed, the source address (SA) is made the retrieval key to select the destination server from the load balancing table 2 shown in Fig.15 (at step T23).

In this example, since the source address (SA) is "2000.12", the server with the destination server address of "S1" is selected.

【0054】

On the other hand, when it is found to be the change notification message of the source address (SA) at step T22, a new entry preparation is executed (at step T24).

In this case, a flow chart of Fig.14B is executed, so that the load balancing table 2 shown in Fig.15 is firstly looked up (at step T231) to determine whether or not an entry already exists (at step T232).

【0055】

As a result, if the entry is not prepared, namely, if the retrieval where an old care-of address (=CoA1) included in the notification message ⑤ is made the retrieval key results in a mishit, an entry is newly prepared.

At this time, the retrieval key to be registered is the new care-of address. However, as the contents of the prepared entry, the table 2 is looked up again with the old care-of address included in the notification message ⑤ as the retrieval key and the resulting old server address obtained is stored as the data of the entry whose retrieval key is the new care-of address as shown in Fig.15 (at steps T233 and T234). This is for selecting the same server even if the same mobile IP terminal moves.

【0056】

Thus, the server address used in connection with the old care-of address is copied to the entry of the new care-of address, so that it is made possible to connect to the same server even if the mobile IP terminal moves.

It is to be noted that in the table 2 of Fig.15, a lifetime is shown. This lifetime indicates the lifetime of the entry, to which a subtraction is processed every fixed period of time to delete the entry when it assumes "0" as a timeout.

【0057】

Therefore, the source addresses (SA's) coexist for a fixed lifetime for the same server address, so that it is not required to prepare a new entry when the mobile IP terminal MN again returns to the care-of address CoA1 (at step T235).

【0058】

It is to be noted that in the embodiment (2) shown in Figs.11 and 12, the load balancer LB performs the change notification request of the care-of address to the home agent HA. However, since the IP terminal itself originally knows the change of the care-of address,

when the load balancer LB requests this IP terminal address to notify upon change of the care-of address in the same way, the same operation as mentioned above is performed.

【0059】

As described above, a load balancer according to the present invention is arranged such that identifying information specific to a mobile IP terminal is extracted from an arrival packet, and a destination server to be connected is determined based on the identifying information. Therefore, at the time of a server access from the mobile IP terminal, a server connection consistency is always maintained.

【0060】

Also, by requesting a home agent or a terminal to notify a change of a care-of address when the care-of address of the terminal has changed and by determining a destination server to be connected by regarding the notified care-of address as identifying information, the server connection consistency can be similarly maintained.